Fundamental structures of dynamic social networks

Sekara, Vedran; Stopczynski, Arkadiusz; Jørgensen, Sune Lehmann

Published in:
Proceedings of the National Academy of Sciences of the United States of America

Link to article, DOI:
10.1073/pnas.1602803113

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Fundamental structures of dynamic social networks

Vedran Sekara[a], Arkadiusz Stopczynski[a,b], and Sune Lehmann[a,c,1]

[a]Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark, [b]Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]The Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark

Edited by Albert-Laszlo Barabasi, Northeastern University, Boston, MA, and accepted by Editorial Board Member Kenneth W. Wachter July 12, 2016 (received for review March 9, 2016)

Social systems are in a constant state of flux, with dynamics spanning from minute-by-minute changes to patterns present on the timescale of years. Accurate models of social dynamics are important for understanding the spreading of influence or diseases, formation of friendships, and the productivity of teams. Although there has been much progress on understanding complex networks over the past decade, little is known about the regularities governing the microdynamics of social networks. Here, we explore the dynamic social network of a densely-connected population of ∼1,000 individuals and their interactions in the network of real-world person-to-person proximity measured via Bluetooth, as well as their telecommunication networks, online social media contacts, geolocation, and demographic data. These high-resolution data allow us to observe social groups directly, rendering community detection unnecessary. Starting from 5-min time slices, we uncover dynamic social structures expressed on multiple timescales. On the hourly timescale, we find that gatherings are fluid, with members coming and going, but organized via a stable core of individuals. Each core represents a social context. Cores exhibit a pattern of recurring meetings across weeks and months, each with varying degrees of regularity. Taken together, these findings provide a powerful simplification of the social network, where cores represent fundamental structures expressed with strong temporal and spatial regularity. Using this framework, we explore the complex interplay between social and geospatial behavior, documenting how the formation of cores is preceded by coordination behavior in the communication networks and demonstrating that social behavior can be predicted with high precision.

complex networks | social systems | human dynamics | computational social science | human mobility

Human societies, their organizations, and communities give rise to complex social dynamics that are challenging to understand, describe, and predict. Recently, network science has provided a powerful mathematical framework for describing the structure and dynamics of social systems (1, 2). With deep roots in traditional sociology (3, 4), a central challenge in the description of social systems is understanding social group behavior. Using empirical data, such groups (or communities) have recently been shown to be highly overlapping and organized in a hierarchical manner (5–8). Without an understanding of the fundamental meso-level structures and regularities governing social systems, modeling behavioral patterns remains a challenge (9).

Although a coherent mathematical framework is not yet in place, existing research suggests that social dynamics are far from random. For example, the existence of strong regularities for individuals in human populations has been well documented within mobility patterns (10, 11), and in social systems, pair-wise interactions show clear patterns occurring at multiple timescales from seconds to months (12). For groups of interacting individuals, however, an understanding of the fundamental structures and their temporal evolution across timescales has proven elusive so far, suggesting a potential for a better understanding as well as models describing important processes such as spreading of influence or diseases, formation of friendships, or productivity of teams.

Our work is based on a longitudinal (36-mo) high-resolution dataset describing a densely connected population of ∼1,000 freshman students at a large European university (13). We consider interactions in the network of physical proximity measured via Bluetooth (*Materials and Methods*), complemented with information from telecommunication networks (phone calls and text messages), online social media (Facebook interactions), as well as geolocation and demographic data.

Until this point, community detection in dynamic networks has required complex mathematical heuristics (14, 15). Here, we show that with high-resolution data describing social interactions, community detection is unnecessary. When single time slices are shorter than the rate at which social gatherings change, communities of individuals can be observed directly and with little ambiguity (Fig. 1*A*). Using a simple matching between time slices, we can infer temporal communities. These dynamic communities offer a powerful simplification of the complex system of social interactions as it develops over time.

In *Results*, we describe these findings in detail. We study the properties of gatherings and cores, show how appearances of social cores are preceded by increased communication among their members, and examine how cores tend to behave as social units. As an application of the framework, we show that the social dynamics of this population are highly predictable.

Results

Social interactions unfold on many timescales, with structures and regularities spanning from minute-by-minute changes to yearly rhythms and beyond, as observed in telecommunication networks (12) or online social networks (16). Despite these inherent dynamics, most of the current understanding of social networks comes from the study of static network topologies, not including temporal aspects (1, 2). Human social communities are inherently overlapping (5, 6, 8, 17), with each and every individual participating in multiple communities. In time-aggregated

Significance

We study the dynamic network of real world person-to-person interactions between approximately 1,000 individuals with 5-min resolution across several months. There is currently no coherent theoretical framework for summarizing the tens of thousands of interactions per day in this complex network, but here we show that at the right temporal resolution, social groups can be identified directly. We outline and validate a framework that enables us to study the statistical properties of individual social events as well as series of meetings across weeks and months. Representing the dynamic network as sequences of such meetings reduces the complexity of the system dramatically. We illustrate the usefulness of the framework by investigating the predictability of human social activity.

Author contributions: V.S., A.S., and S.L. designed research; V.S. and A.S. performed research; V.S., A.S., and S.L. analyzed data; and V.S., A.S., and S.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.-L.B. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

See Commentary on page 9961.

[1]To whom correspondence should be addressed. Email: sljo@dtu.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1602803113/-/DCSupplemental.

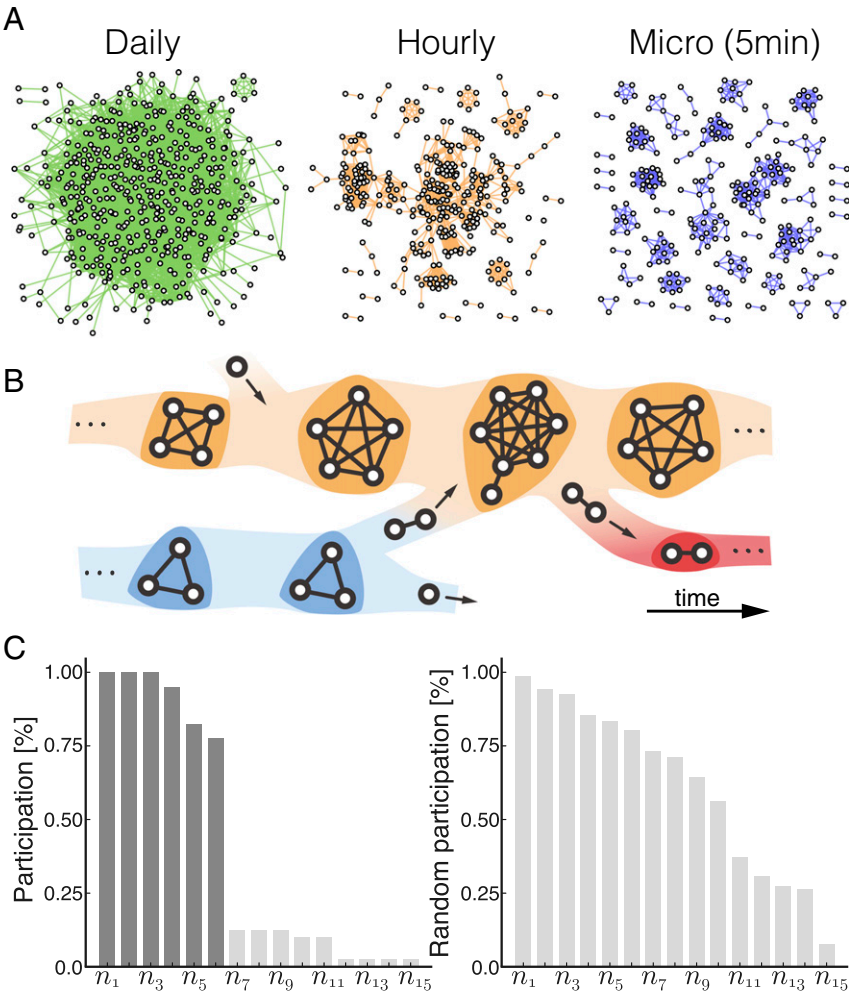

Fig. 1. Properties of gatherings. (*A*) The social network at different timescales. The network formed by face-to-face meetings within 1-d (green), 60-min (orange), and 5-min (blue) temporal aggregation. In the 5-min time slices, groups are directly observable without much ambiguity, but the overlap between groups increases as time is aggregated across bins. (*B*) Illustration of gathering dynamics. Gatherings evolve gradually with members flowing in and out of social contexts. (*C*) Extracting cores from participation profiles. Dark-gray bars denote nodes with participation levels above the maximal gap. Ordered participation profile for empirical data (*Left*), as a null model we use participation profiles generated from a uniform random distribution (*Right*).

networks, this structural property results in communities with more outgoing than incoming edges, the "hairball" structure visualized in Fig. 1*A* (green network). With minute-by-minute observation intervals of this network, individuals are constrained to a single community per time slice, and we can observe the basic structural elements directly and with little ambiguity, as illustrated in Fig. 1*A* (blue network). More generally, when time slices have a duration that is shorter than the rate with which the group composition changes, gatherings of individuals can be observed directly, rendering traditional community detection algorithms redundant. Matching connected components across time slices, we can study the temporal development of social "gatherings." In our dataset, gatherings are typically together for a few hours (full statistics are available in *SI Appendix*, *S3: Gatherings*, *S3.3: Gathering Statistics*) and can be thought of as instances of social structures ("cores") that occur repeatedly over days, weeks, and months.

Gatherings. In the physical proximity network, meetings require that all members be present at the same time and that they are in close physical proximity. These properties of physical meetings imply that gatherings can be directly identified as graph components within each 5-min time slice. Therefore, we are able to use a simple hierarchical clustering to match groups across time slices (*SI Appendix*, *S3: Gatherings*, *S3.1: Detecting Gatherings*), using a matching strategy similar to ref. 18.

A node may be part of only a single gathering per time step, but, over time, individuals flow in and out of social gatherings, shifting their affiliation and forming new gatherings, as illustrated in Fig. 1*B*. The gatherings display broad distributions in both size and duration, capturing meetings ranging from small cliques to large aggregations and from short interactions on the order of minutes to prolonged encounters lasting many hours, covering a wide range of meeting types. Gatherings are defined for any number of nodes greater than one, but because we are interested in group dynamics, we only discuss gatherings of size three or greater in the statistics reported here. Because our cohort consists of university students, an important inhomogeneity in the data is between "work" activities that take place on campus (including scheduled classes) and "recreational" activities that take place off campus. Using global positioning system (GPS) information, we find that 42% of gathering take place on-campus (work) and 58% take place off-campus (recreation). Comparing work/recreation statistics, recreational gatherings tend to be smaller but last considerably longer, illustrating that the context of meetings can influence their properties (see *SI Appendix*, *S3: Gatherings* for full statistics).

The fluid behavior illustrated in Fig. 1*B* results in "soft" gathering boundaries, with some members participating for the total duration of the gathering and others participating only briefly. Peripheral members can be acquaintances briefly interacting with members of the social core but can also be spurious connections in the data: nearby strangers eliciting a Bluetooth measurement not corresponding to a social connection. Despite these soft boundaries, we find that gatherings are characterized by a stable core of individuals that are present during the majority of each meeting (*SI Appendix*, *S4: Cores*, *S4.1: Extracting Cores*). This stable core is expressed in participation profiles as illustrated in Fig. 1*C*. Here, a participation profile is the sorted fraction of time each node has participated in a particular social

context, normalized by its total lifetime. A pronounced core structure implies a gap in the participation profile, separating the core members (Fig. 1*C*, *Left*, dark-gray bars) from the peripheral nodes (Fig. 1*C*, *Left*, light-gray bars). We test whether the gap is statistically significant by comparing to an ensemble of profiles generated by a random process (Fig. 1*C*, *Right*), where a random participation level between 0 and 1 is assigned to each node from a uniform distribution. Based on the ensemble, we estimate the average expected gap size and deviation. If the gap observed in the empirical data is greater than the average null-model gap μ_{random} plus 1 SD σ_{random}, we accept the core as significant. According to this criterion, we find that 7,146 out of the 7,320 (97.6%) inferred groups display a pronounced core structure.

Cores. A gathering represents the time evolution of a single meeting between a group of individuals. In most cases, however, gatherings are an instance of a lasting social context (e.g., a group of friends or classmates), and we observe the same nodes participating in subsequent gatherings occurring repeatedly over the following days, weeks, and months. We call the social structures corresponding to all gatherings of the same set of individuals "cores." We argue below that cores represent a fundamental structure expressed in dynamic social networks with strong temporal and spatial regularity. Below we discuss cores. How cores are identified, and how cores can be used to quantify the regularity of social interactions, as well as predict future behavior of individuals in our dataset.

In this population, the number of appearances per core is a heavy-tailed distribution; most cores appear only a few times, whereas the most active cores can appear multiple times per day over the full observation period (*SI Appendix*, *S4: Cores*, *S4.2: Core Statistics*). Here, we focus on the temporal patterns of recurring gatherings, restricting our analysis to cores of size three or greater that are observed more than once per month, on average. The split between activities that occur on campus is important for cores as well as gatherings, and Fig. 2*A* shows the difference of how individuals engage and spend time in different social contexts ("work cores"—primarily observed on campus—and "recreational cores"—primarily observed elsewhere). Although the distribution of number of recreational cores per person is broad, most participants are part of only one or two recreational cores (Fig. 2*A*, *Upper*). The distribution of work cores per person is localized with an average of 2.74 ± 1.85 work cores per person (presumably corresponding to classes and study groups).

We find that cores leave traces in other data channels that emphasize the differences between work and recreational meetings. One such trace is coordination behavior, which we can explore by investigating how call and text-message activity increases in the time leading up to a meeting. We define a core's level of coordination c_t at time t as the average increase of activity of its members before a meeting. Specifically, we set $c_t = 1/N\sum_{n=1}^{N} a_t^n/\widetilde{a_t^n}$, where N is the number of participants and a_t^n is the individual activity of person n in time-bin t (indexing the hours-of-the-week), compared with an individual baseline, which is simply the person's average activity $\widetilde{a_t^n}$ in that hour of the week; to generate the curves in Fig. 2, we then average over cores. We observe clear evidence of coordination before meetings, with a stronger effect during weekends (Fig. 2*B*, *Upper*). This result explicitly shows the extra coordination cost associated with non–schedule-driven interactions. Conversely, the coordination cost per participant does not depend on the size of the gathering in question (Fig. 2*B*, *Lower*), suggesting a social optimization process takes place [because the number of social connections grows as $n(n-1)/2$ for a group of n participants].

Modeling the Social Network. To place the significance of the basic statistics regarding gatherings and cores in perspective, we create a null model for the full network dynamics (see *SI Appendix*, *S7: A Dynamic Random Geometric Graph Model* for full details). We model each time slice as a random geometric graph (RGG) and generate dynamics by representing each node as a random walker, similar to models known from the literature (19). For reasonable parameter values, we find that this "dynamic RGG model" creates gathering-like components and is able to recreate qualitative features of the distribution of both gathering sizes and lifetimes. Thus, as suggested in the literature (19), we are in fact able to model key statistical features of gathering occurrences across a single day, using a very simple model.

The fundamental difference between our empirical system and the dynamic RGG model arises because the model does not capture correlations between groups of nodes. This inability to capture correlations means that the dynamic RGG model does not generate recurring gatherings and is unable to model dynamics across weeks and months. The fact that gatherings in the dynamic RGG model only occur once means that empirical statistics on the level of cores are not reproduced by the model.

Cores Are Social Units. It is well established that two individuals who share a social tie have correlated mobility patterns (20, 21). Because interactions with social ties tend to happen at irregular intervals (22), using this correlation for location prediction has proven difficult. Expressed differently, we know that the location of person A is going to be predictive of person B's location (and

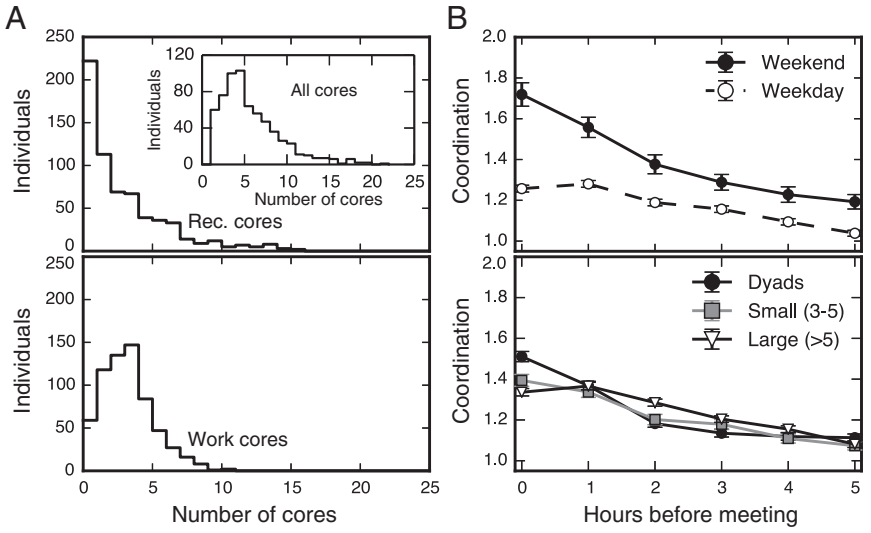

Fig. 2. Cores summarize social contexts for individuals. (*A*) The distributions of work and recreational (Rec.) core memberships; the upper graph shows that individuals typically participate in only one or two recreational contexts, although the tail of the distribution contains individuals with more gregarious behavior. (*A*, *Inset*) The distribution of participation in cores overall (both work and recreation). (*B*) Coordination before meetings defined as increase of phone calls and messages occurring within hourly time bins before the meeting relative to a null model based on the average hourly telecommunication behavior for each participant. More coordination is required to organize meetings during weekends than during weekdays. Larger meetings do not require additional coordination per participant.

vice versa) when they meet, but we rarely know at which point in time A and B will meet.

As we have defined them, cores represent meetings between $n > 2$ individuals, characterized by the fact that all n members are usually present when the core is active. Therefore, an observation of an incomplete set of core members implies that the remaining members will arrive shortly, a fact that can be used for prediction. We call this property of cores the "social unit" attribute.

We illustrate the social unit attribute using cores of size three. Given that two members of a core are observed, we measure the probability that the remaining member will arrive within one hour. To avoid testing on scheduled meetings, we only consider weekends and weekday evenings and nights (6:00 PM to 8:00 AM), where meetings are not driven by an academic schedule. Furthermore, we test on a month of data that has not been used for identifying cores. We compare the behavior of actual cores to two null models, both based on an aggregated graph of daily interactions.

The first null model (random) simply illustrates that our results are not driven by spurious colocations. In the random null model, we create an aggregated daily graph for each day in the observation period. We then choose a random day and pick three individuals at random. If two of those individuals appear in the same location, we measure the probability of the third node arriving within one hour (including the reference group in the statistics only if at least two of its nodes are colocated during the day; *SI Appendix*, *S6: Social Prediction*). This situation is illustrated using blue nodes in Fig. 3*A*. The random null model shows no predictive power (Fig. 3*B*).

For the second null model [breadth first search (BFS)], we test the effect of pairwise friendships within groups of three. Here, we form reference groups using a BFS strategy. We choose a random day and based on a randomly chosen seed node, we perform BFS steps until the neighborhood set is large enough to form a group of size three; then, we create a reference group based on the seed and two random nodes from the neighborhood. This situation is illustrated in Fig. 3*A*, where the BFS starts from a yellow seed node expanding to two randomly chosen neighbors marked in green. We then choose situations where two nodes from this set are subsequently colocated and test how often the third BFS reference-group member arrives within the next 60 min (*SI Appendix*, *S6: Social Prediction*). The BFS model implies at least pairwise relationships between the (yellow) seed node and the two other (green) group members.

Comparing the performance of cores with the two null models (Fig. 3*B*), we find support for the social unit attribute. The random null model clearly shows that spurious connections do not carry a signal, and the BFS model illustrates that even if all three nodes form a connected subgraph, those subgraphs have a less than 10% probability of the third member arriving within the hour. In contrast, the cores capture arrivals of the remaining member in close to 50% of cases despite cores having soft boundaries. These observations provide support for, and quantify, the social unit attribute. Cores require all members to be present.

Ego Perspective: Core Participation Forms Social Trajectories. From the perspective of an individual, we find the situation displayed in Fig. 3*C*, where each ego is involved in a number of overlapping and nested cores; the distribution (number of memberships) is shown in Fig. 2*A*, *Inset*. Fig. 3*D* shows when cores are activated across our observation period from ultimo January to ultimo April, with time running horizontally and each row/color corresponding to a core, with colors matching Fig. 3*C*; we call the core instantiation profile an individual's "social trajectory."

Replacing the complex temporal dynamics of the social network with the sequence of participations in the core instantiation profile offers a powerful simplification of the complexity of a dynamic social network. Because each core represents a social context the set of an individual's cores provides a finite set of states (a "vocabulary") for quantifying their social life.

The social trajectory shown in Fig. 3*D* also provides an important connection to research within human mobility. Based on

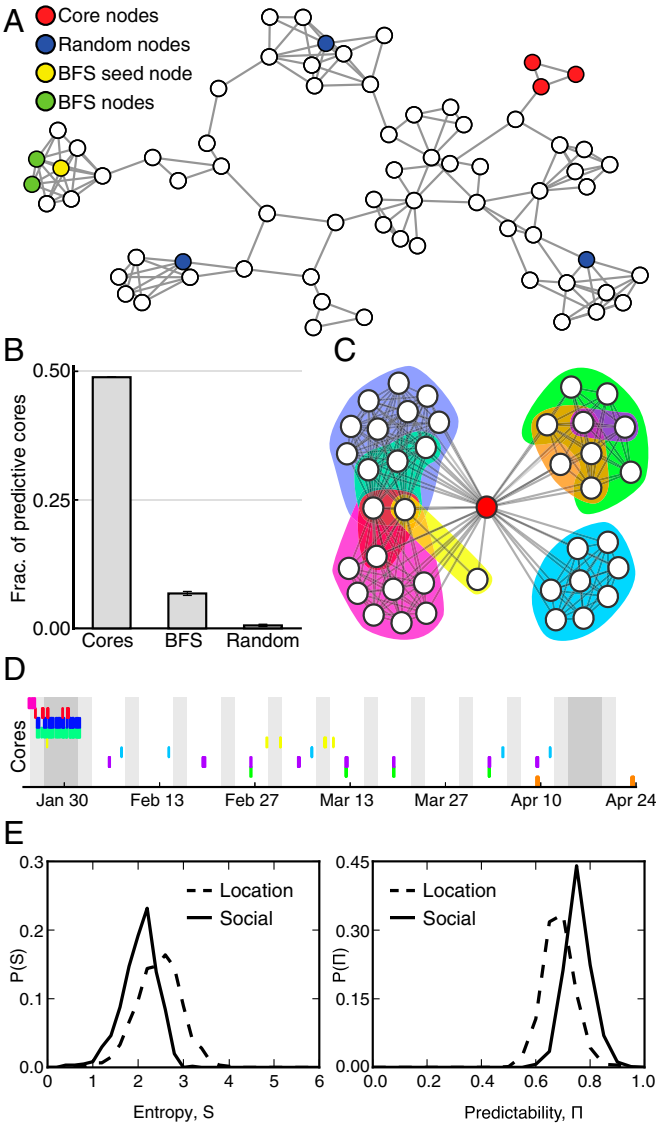

Fig. 3. Cores predict social behavior. (*A*) Example of an aggregated graph of daily interactions, illustrating cores and construction of null models. Blue nodes correspond to the random null model, yellow/green nodes illustrate the BFS null model, and red nodes show a three-core. See main text for full details. (*B*) Percentage of predictive groups in each category: Cores, BFS reference-groups, and random groups. Reference model error bars are calculated across $n = 100$ independent trials. (*C*) Example ego view of communities; we observe overlapping and nested structures. (*D*) The temporal instantiations of the cores in *C*. Time runs horizontally and each row corresponds to cores (colors matching *C*). (*E*) The probability distributions P(S) and P(Π) of time-correlated entropy S and predictability Π for social and location patterns, respectively. We find that overall social patterns tend toward lower entropy than geospatial traces, resulting in higher predictability.

mobility data, it has recently been shown that human mobility patterns are regular, and that given a sequence of location-observations, a person's geographical location in the next time bin can be predicted with high accuracy (an average of 93% of the time) (11). Interestingly, the problem of predicting the next instantiation of a social core is equivalent to predicting the next location in a sequence of locations, once we summarize the dynamic network of social interactions using the social trajectory formulation. In the analogy, a social core corresponds to a location in physical space, and instantiating a gathering corresponds to visiting that location. This implies that we can use the methods developed in ref. 11 to estimate an upper bound for the predictability of

social trajectories. The upper limit of predictability is derived from the amount of repetition (routine) encoded in a sequence of observations.

The central property needed to estimate the bound on predictability is the time-correlated entropy S, which quantifies the amount of uncertainty within a data sequence, accounting for frequency and ordering of states. The upper bound on predictability, Π, can then be determined by applying a limiting case of Fano's inequality (23) (*SI Appendix*, *S5: Predicting Behavior from Routine*).

Because we have access to participants' location traces as well as their networks, we can calculate the upper bound on predictability for both social trajectories and location traces across the population. To incorporate the full complexity of social interactions in the calculation of the time-correlated entropy, we include cores with any number of appearances, as well as cores consisting of two nodes in these calculations. In Fig. 3*E*, we display the distributions of time-correlated entropy and corresponding upper bound on predictability for the population. Comparing the two distributions, we find that the typical social behavior is characterized by lower entropy and thus higher predictability than the typical mobility behavior. In the dynamic RGG model, which we use as a null model for the full network dynamics, the temporal entropy does not converge; this absence of convergence implies that there is no social predictability in this model (*SI Appendix*, *S7: A Dynamic Random Geometric Graph Model*).

Comparing the empirical mobility behavior to the literature (11, 24), we find a lower average predictability limit of approximately 80% (Fig. 3*A*). The fact that the location predictability is somewhat lower than the 93% previously reported (11) is connected to a number of factors: the nonrepresentative sociodemographics of the study population may play a role, the GPS location data used here have significantly higher spatial resolution than the cell towers used in previous work (11, 24)—a fact that is known to decrease the estimated predictability (25)—and we focus on predicting the next location in a sequence of locations rather than predicting the person's location in the next time-bin—a more challenging prediction task (*SI Appendix*, *S5: Predicting Behavior from Routine*, *S5.1: Comparing with Previous Studies*).

The overall level of social and geospatial predictability is not correlated for individuals ($P = 0.85$; *SI Appendix*, *S5: Predicting Behavior from Routine*, *S5.4: Temporal Aspect of Predictability*). Despite this lack of correlation, we know that the predictability encoded in social or geospatial trajectories is closely related to daily and weekly schedules (26). A deeper exploration of the connection between social and geospatial behavior across the week provides an example of how the social trajectories enable a joint sociospatial description of an entire population.

First, we consider the typical behavior by calculating the uncorrelated entropy (*Materials and Methods*) for the social and geospatial trajectories independently. As a measure of disorder, the entropy quantifies how unpredictable a pattern is: the higher the entropy, the more users tend to distribute their time between many distinct states. Fig. 4*A* shows the entropy averaged over the population for each day of the week, in 8-h bins: "nights" running from midnight to 8:00 AM, "days" spanning 8:00 AM to 4:00 PM, and "evenings" from 4:00 PM to midnight. Light colors indicate high entropy (complex behavior) and dark colors low entropy (simple patterns).

The geospatial behavior (Fig. 4*A*, *Upper*) is characterized by low entropy on weeknights, essentially corresponding to sleeping in one or two locations. The entropy is high during weekdays and assumes midrange values during most evenings with a strong exception on Friday and Saturday nights, which are characterized by the highest entropy of the entire week. These "party nights," are consistently the time bins with highest average geospatial entropy, corresponding to exploration behavior.

The social behavior (Fig. 4*A*, *Lower*) is similar to the geospatial behavior across most of the week: predictable nights, varied days, and evenings somewhere in between. On Friday and Saturday night, however, the social trajectories display behavior that is significantly different from the collective pattern arising from the geospatial trajectories. In these time bins, when the study participants are most exploratory in a geospatial sense, the participants appear to be highly conservative in a social sense, displaying simpler and more predictable social behavior. This finding is consistent with Fig. 2*A*, which shows that a large majority of participants focus their off-campus life on a small number (one or two) of social cores.

The observations above suggest that during the week, the population is characterized by a pattern of the same people meeting in the same places, with exploration peaking on Friday nights and weekends. On Friday night and weekends when the population is most unpredictable in its geospatial behavior, the same individuals tend to be highly predictable in their social behavior, exploring a range of locations, but always with the same core of friends.

In interpreting this result, it is important to realize that we only observe the social interactions among participants of the experiment. Whereas the geospatial data stream is sampled evenly over the observation period, the social stream has a potential bias (e.g., it is possible to go out with nonuniversity friends on Fridays and weekends). We can address this caveat by considering the behavior of cores across the week. In Fig. 4*B*, we consider the behavior of cores across time, displaying the uncorrelated location entropy of "core location histories," i.e., the sequence of locations visited by each core (averaged and binned as above). In interpreting the absolute values of entropy in Fig. 4*B*, note that cores typically only meet in a few locations; thus, in turn, cores have fewer location states than individuals, resulting in smaller average values for the entropy.

Fig. 4*B* shows directly that cores do display a distinct exploratory behavior on Friday nights (and weekends more generally).

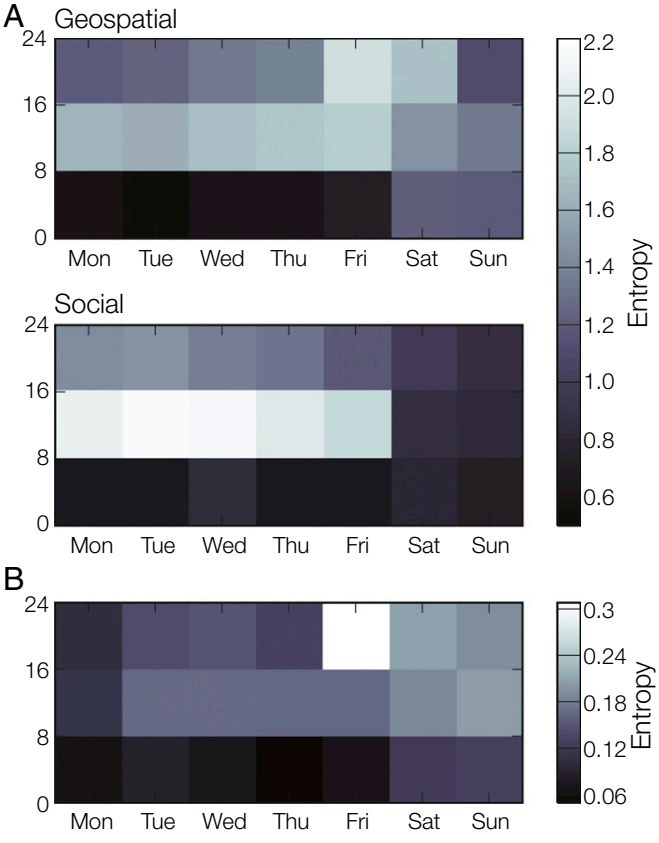

Fig. 4. Temporal aspects of geospatial and social predictability. (*A*) Uncorrelated entropy respectively for location (*Upper*) and social (*Lower*) states, averaged across all individuals, showing how unpredictable individuals are within 8-h bins over the course of a week. (*B*) Geospatial entropy of cores, indicating when cores (i.e., individuals together in a social setting) explore new geospatial locations.

The fact that geospatial exploration occurs as part of a social group, but constrained to certain time bins, reveals a complex interplay between time, location, and social context, and supports the hypothesis that at times, when humans are most unpredictable in the geospatial domain, they display predictable social behavior.

Discussion

The freshman participants of this study are not a representative sample of society as a whole, and we expect that some aspects of the findings presented here reflect that this is a particularly youthful demographic sample (e.g., geospatial entropy peaking on Friday and weekend nights). Although the population studied here is not representative of society as a whole, we argue that the methods developed, as well as many of our findings do generalize to more representative populations. Below we summarize the most important of those findings.

It is often the case that the mathematical description of networks grows much more complex once the temporal dimension is included (9), with even basic network properties, such as degree, clustering coefficient, or centrality having multiple competing generalizations to the temporal domain. Here, we observe the opposite effect that additional temporal information can simplify the description. By observing the network at high temporal resolution, we can directly observe gatherings, a phenomenon that is obscured in time-aggregated networks (8). In another context (17), data on empirical network flow has been used as extra information to uncover overlapping communities. Here, a simple matching across time slices reveals dynamically changing gatherings with stable cores that provide a strong simplification of the social dynamics. These cores leave traces in other data channels, for example, eliciting coordination in the telecommunication network. The cores are fundamental units in the sense that they require all members to be present, a fact that we can use to predict the future location of other (nonpresent) members.

Connecting our findings to the literature of dynamic community detection, we note that many elegant methods exist in the literature that would allow us to detect gatherings in a daily network (6, 14, 15, 27–28), but here we have chosen to use a simple matching of graph components to underline that social structures in individual snapshots are so obvious that these sophisticated methods are unnecessary. None of the algorithms cited above, however, are designed for detecting communities at multiple temporal scales; therefore, those methods are unable to identify cores and help us understand the patterns of cores over time.

Using an individual's cores as a set of social states, we can define a social trajectory that simplifies the description of the individual's social activity, allowing for an exploration of the predictability encoded in social routine. We find that, for individuals in our population, times of exploration in the geospatial domain are connected to a small subset of social circles, suggesting a general hypothesis that when humans are most unpredictable with respect to location we are most predictable with respect to our social context.

In summary, our work provides a quantitative look at the rich long-term patterns encoded in the microdynamics of a large system of interacting individuals, characterized by a high degree of order and predictability. The work presented here provides a framework for describing human behavior and hints at the promise of our approach. Although we have focused on predictability, we expect that our work will support better modeling of a multitude of processes in social systems, from epidemiology and social contagion to urban planning, organizational research, and public health.

Materials and Methods

The Dataset. The proximity interaction network of participants from the Copenhagen Networks Study was based on Bluetooth scans, collected using smartphones. Physical proximity measured via Bluetooth corresponds to distance between 0 and ~10 m, depending on environmental conditions. Because false positives (reported observations of users not actually present) in Bluetooth scans are unlikely, we use a symmetrized (undirected) network of interactions.

Informed Consent. Data collection was approved by the Danish Data Protection Agency, and informed consent has been obtained for all study all participants.

Definitions of Entropy. For an individual i's sequence of states, we define the uncorrelated entropy $S_i^{unc} = -\sum_j^{N_i} p_j \log_2 p_j$, where p_j is the probability of observing state j, and time-correlated entropy $S_i^{temp} = -\sum_{T_i' \subset T_i} p(T_i') \log_2 [p(T_i')]$, where $p(T_i')$ is the probability of observing the subsequence T_i' in the trajectory T_i.

ACKNOWLEDGMENTS. We thank L. K. Hansen, P. Sapiezynski, A. Cuttone, D. Wind, J. E. Larsen, B. S. Jensen, D. D. Lassen, M. A. Pedersen, A. Blok, T. B. Jorgensen, and Y. Y. Ahn for invaluable discussions and comments on the manuscript and R. Gatej for technical assistance. This work was supported by the Villum Foundation [Young Investigator Program "High Resolution Networks" grant (to S.L.)] and the University of Copenhagen (UCPH Excellence Programme for Interdisciplinary Research "Social Fabric" grant).

1. Easley D, Kleinberg J (2010) *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge Univ Press, Cambridge, UK).
2. Newman M (2010) *Networks: An Introduction* (Oxford Univ Press, Oxford).
3. Simmel G (1950) Quantitative aspects of the group. *The Sociology of Georg Simmel* (Free Press, Glencoe, IL), pp 87–177.
4. Goffman E (2005) *Interaction Ritual: Essays in Face to Face Behavior* (AldineTransaction, Piscataway, NJ).
5. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043):814–818.
6. Palla G, Barabási AL, Vicsek T (2007) Quantifying social group evolution. *Nature* 446(7136):664–667.
7. Clauset A, Moore C, Newman ME (2008) Hierarchical structure and the prediction of missing links in networks. *Nature* 453(7191):98–101.
8. Ahn YY, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. *Nature* 466(7307):761–764.
9. Holme P, Saramäki J (2012) Temporal networks. *Phys Rep* 519(3):97–125.
10. González MC, Hidalgo CA, Barabási AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782.
11. Song C, Qu Z, Blumm N, Barabási AL (2010) Limits of predictability in human mobility. *Science* 327(5968):1018–1021.
12. Saramäki J, Moro E (2015) From seconds to months: An overview of multi-scale dynamics of mobile telephone calls. *Eur Phys J B* 88(6):164.
13. Stopczynski A, et al. (2014) Measuring large-scale social networks with high resolution. *PLoS One* 9(4):e95978.
14. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328(5980):876–878.
15. Gauvin L, Panisson A, Cattuto C (2014) Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS One* 9(1):e86028.
16. Golder SA, Macy MW (2011) Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333(6051):1878–1881.
17. Rosvall M, Esquivel AV, Lancichinetti A, West JD, Lambiotte R (2014) Memory in network flows and its effects on spreading dynamics and community detection. *Nat Commun* 5:4630.
18. Greene D, Doyle D, Cunningham P (2010) Tracking the evolution of communities in dynamic social networks. *Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining* (IEEE Computer Society, Washington, DC), pp 176–183.
19. Starnini M, Baronchelli A, Pastor-Satorras R (2013) Modeling human dynamics of face-to-face interaction networks. *Phys Rev Lett* 110(16):168701.
20. Wang D, Pedreschi D, Song C, Giannotti F, Barabási AL (2011) Human mobility, social ties, and link prediction. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York), pp 1100–1108.
21. De Domenico M, Lima A, Musolesi M (2013) Interdependence and predictability of human mobility and social interactions. *Pervasive Mob Comput* 9:798–807.
22. Miritello G (2013) *Temporal Patterns of Communication in Social Networks* (Springer Science & Business Media, Berlin).
23. Fano RM (1961) *Transmission of Information: A Statistical Theory of Communications* (MIT Press, Cambridge, MA).
24. Lu X, Wetter E, Bharti N, Tatem AJ, Bengtsson L (2013) Approaching the limit of predictability in human mobility. *Sci Rep* 3:2923.
25. Lin M, Hsu WJ, Lee ZQ (2012) *Predictability of Individuals' Mobility with High-Resolution Positioning Data* (Association for Computing Machinery, New York), pp 381–390.
26. McInerney J, Stein S, Rogers A, Jennings NR, Exploring periods of low predictability in daily life mobility, Nokia Mobile Data Challenge 2012 Workshop, June 18, 2012, Newcastle, UK (University of Southhampton, Southhampton, UK).
27. Rosvall M, Bergstrom CT (2010) Mapping change in large networks. *PLoS One* 5(1):e8694.
28. De Domenico M, Lancichinetti A, Arenas A, Rosvall M (2015) Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys Rev X* 5:011027.
29. Salnikov V, Schaub MT, Lambiotte R (2016) Using higher-order Markov models to reveal flow-based communities in networks. *Sci Rep* 6:23194.